



RESEARCH DEPARTMENT



REPORT

An evaluation of the use of microphone arrays for studio recording

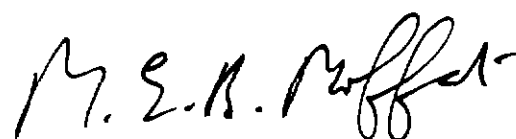
P.S. Gaskell, M.A., C.Eng., M.I.E.E., and K.A. Bradbury, B.A.

**AN EVALUATION OF THE USE OF MICROPHONE ARRAYS
FOR STUDIO RECORDING**
P.S. Gaskell, M.A., C.Eng, M.I.E.E., and K.A. Bradbury, B.A.

Summary

This Report assesses the feasibility of using arrays of microphones as a studio recording technique in those situations where it is difficult to site the microphone close to the source. Theoretical studies and practical tests have shown that the reverberant field, present in all studios, imposes severe demands on the performance of the array. It is estimated that the minimum number of microphones needed for a practical array is of the order of 100. The array would have to be constructed to stringent practical tolerances and would need considerable signal processing hardware for co-phasing and beam steering. This technique therefore is unlikely to be appropriate for studio recording except in special circumstances.

Issued under the Authority of



Head of Research Department

**Research Department, Engineering Division,
BRITISH BROADCASTING CORPORATION**

August, 1985
(PH-273)

This Report may not be reproduced or disclosed to a third party in any form without the written permission of the British Broadcasting Corporation.

It uses SI units in accordance with B.S. document PD 5686.

AN EVALUATION OF THE USE OF MICROPHONE ARRAYS FOR STUDIO RECORDING

P.S. Gaskell, M.A., C.Eng, M.I.E.E., and K.A. Bradbury, B.A.

Section	Title	Page
	Summary.....	Title Page
1.	Introduction.....	1
2.	Theoretical prediction of array performance	1
	2.1. General.....	1
	2.2. The response to direct sound	2
	2.3. Reverberant sound	3
	2.3.1. Single microphone response	3
	2.3.2. Array response	3
	2.3.3. Distance of equal response	3
	2.4. The combined direct and reverberant sounds	4
	2.4.1. Presence	4
	2.4.2. Discrimination between sources	4
3.	Computer predictions	5
	3.1. Design of arrays	5
	3.2. Array configurations	5
	3.3. Simulation results	6
4.	Practical tests	7
	4.1. Array design	7
	4.2. Experimental arrangements.....	8
	4.3. Component amplitude and phase responses	8
5.	Results of the practical tests.....	9
	5.1. Direct sound in free-field room	9
	5.1.1. Tone tests.....	9
	5.1.2. Speech tests	10
	5.1.3. Microphone noise	11
	5.2. Reverberant sound in reverberation room	11
	5.2.1. Noise tests.....	11
	5.2.2. Speech tests	13
	5.3. Performance in a "small TV studio"	13
	5.3.1. Noise tests.....	13
	5.3.2. Speech tests	14
6.	Discussion	14
	6.1. Practical tests	14
	6.2. Characteristics of practical arrays	15
7.	Conclusions	15
8.	References	16
	Appendix 1	16

AN EVALUATION OF THE USE OF MICROPHONE ARRAYS FOR STUDIO RECORDING

P.S. Gaskell, M.A., C.Eng, M.I.E.E., and K.A. Bradbury, B.A.

1. Introduction

Virtually all programmes require the sound from a source to be picked up in the presence of other unwanted sounds. Normally, this is done by positioning a microphone close to the wanted source so that the signal levels from that source far exceed those from any other. Sometimes, however, this is impracticable; it may be physically impossible to position a microphone close to the source, (as for example in television drama where the microphones should not be visible) or there may be so many optional sources (as in an audience participation show), that it would be expensive and impractical to provide each of them with a microphone. In these situations, the concept of a highly directional steerable microphone, which could be positioned some distance away and yet still provide the same sound quality as a microphone placed close to the source, is attractive.

It has been suggested that microphone arrays could provide the necessary highly directional beams. A fixed array would be suspended above the studio floor and the wanted source would be selected by suitable post-processing of the individual microphone signals. It is probable that such an array would find greatest application for speech particularly as the performance is likely to be poor at the edges of the audio band.

Array techniques are well established for r.f. and sonar but not for airborne sound. Aerial arrays have long been in use as a means of providing highly directional r.f. beams for radiation or reception. The same approach has been exploited to an advanced degree in sonar.¹ Very large arrays (up to 1,000 transducers) have been constructed to give a precise and discriminating underwater detection method. A wide bandwidth is covered (0.1–5.0 kHz) and beam forming is achieved by time delaying the respective element contributions so that they are co-phased for the wanted source direction.

Sonar and air-borne arrays have much in common in so far as they operate at similar wavelengths and cover a broad band of frequencies. However, there are two significant distinctions. The first is that whereas underwater, the sound source is normally in the far-field and its position is measured

in terms of azimuth only, the source of a microphone array is in the near-field for much of the audio frequency band, and the distance between source and microphone is an important factor. The second is that in the ocean, reverberant sound normally has little effect, whereas in a studio, the reverberant sound is often a critical factor determining the overall selectivity of the array.

This Report considers these differences in detail and discusses their importance in the design of microphone arrays. It then presents the results of a computer simulation and of experimental assessments of some simple arrays.

2. Theoretical prediction of array performance

2.1. General

The sound at any point in a room is a combination of the direct sound from a source and the sounds from that source reflected off the surfaces in the room. The reflected sounds constitute the reverberant component of the sound and, in an acoustically treated studio, it tends to be diffuse and isotropic.

When microphones are used to record individual sources, they may be regarded as fulfilling two principal aims. The first is to pick up the sound of a source with the appropriate balance of the direct and reflected (or reverberant) sounds. The degree to which the sound appears to be close to the listener (known as 'presence') is principally determined by the relative attenuation of the reverberant sounds. An individual microphone achieves this through a combination of its directional characteristics (cardioid, hyper-cardioid etc.) and its proximity to the source (to accentuate the direct sound). Microphone arrays are further from the source and attenuate the reverberant sound by cancellation and summation methods. The second aim is to reject sounds from unwanted sources. Individual microphones again rely on their directionality and placement to achieve this. An array on the other hand, being further from the source, has to form a directional beam to "spot-light" the wanted source and to reject the direct and reverberant sounds from unwanted sources.

The following sub-sections predict an array's response to direct sound only, to reverberant sound only, and to combinations of the two. For simplicity, figures and formulae assume wanted and unwanted sources of equal intensity.

2.2. The response to direct sound

The principles of array beam forming are well known. A planar array forms a broadside far-field beam according to Fraunhofer diffraction theory. The far-field amplitude radiation pattern of a regular, closely spaced, evenly weighted, linear array varies with angle θ from the axis as $\sin x/x$ (the sinc function) where $x = (\pi D \sin \theta)/\lambda$, D is the aperture or width of the array and λ is the wavelength. The angular beam-width or resolution is λ/D . The beam can be steered to directions other than broadside by inserting appropriate delays in the output of each element.

The distance between the elements of an array must be carefully chosen. When the spacing is $\ll \lambda/2$, the beam shape corresponds to $\text{sinc } x$. As the spacing increases, additional $\text{sinc } x$ spectra known as "grating lobes" appear, centred at angles

$$\theta_2 = \arcsin \frac{n\lambda}{d}$$

where d = inter-element spacing and n is an integer. When $d > \lambda$, grating lobes appear at angles $\theta_2 < \pi/2$ and their side-lobes interfere with the main beam pattern. The grating lobes are clearly undesirable since they have the effect of creating multiple beams.

For a microphone array in an enclosure, the source is in the near-field for much of the audio frequency band and the beam-shape deviates from the Fraunhofer pattern. However, the considerations of array aperture and inter-element spacing still broadly apply. As a rule of thumb, the near-field may be considered to be at distances less than D^2/λ . D might typically be 3 m, so D^2/λ is then 26 m at 1 kHz and 2.6 m at 100 Hz. It is most likely that a microphone array would be placed within 5 m of the source and would be in the near-field for all but the lowest frequencies. It is important therefore that the distance of the source from the array is included, as well as its azimuth, when analysing the response of an array. Expressions taking both factors into account were derived to enable accurate computer predictions to be made.

The sound pressure at a distance r from an omnidirectional source is

$$\text{Re} \left\{ \frac{A}{r} \exp j(\omega t - kr) \right\}$$

where A = constant

ω = angular frequency

k = wave number, $2\pi/\lambda$

and "Re" denotes "real part of".

If the array elements are omnidirectional, the sum of the element contributions is

$$\text{Re} \left[A \exp(j\omega t) \sum_{i=1}^N \frac{1}{r_i} \exp(-jkr_i) \right]$$

and its r.m.s. value is

$$\frac{A}{\sqrt{2}} \left| \sum_{i=1}^N \frac{1}{r_i} \exp(-jkr_i) \right|$$

(In subsequent expressions, the common constant term $A/\sqrt{2}$ is omitted for clarity.)

If the distance of the furthest element from the source is $r_{\max}$, then co-phasing is achieved by multiplying the i th element by

$$\exp\{-jk(r_{\max} - r_i)\}$$

This constitutes a constant delay of $(r_{\max} - r_i)/c$. The array beam is now focussed at the source position and the response is

$$\begin{aligned} & \left| \sum_{i=1}^N \frac{1}{r_i} \exp(-jkr_{\max}) \right| \\ &= \sum_{i=1}^N \frac{1}{r_i} \end{aligned} \quad \dots (1)$$

The response to other source positions is

$$\left| \sum_{i=1}^N \frac{1}{r_{io}} \exp(-jkr_{io}) \exp\{-jk(r_{\max} - r_{iw})\} \right| \quad \dots (2)$$

where r_{io} = distance of the i th microphone from the "other" source position, and r_{iw} = distance of the i th microphone from the wanted source position.

The array beam-shape can be tailored by weighting the individual element contributions by a gain factor. In particular, the $\text{sinc } x$ radiation pattern given by a square aperture has a high level of side-lobes. These can be suppressed at the expense of a broader main beam by using, for example, a Hanning window. For the array tests, the window used had a $(0.6 + 0.4 \cos \theta)$ profile where $(-\pi \leq \theta \leq \pi)$ and $\pm \pi$ correspond to the edges of the array.

2.3. Reverberant sound

The selectivity of a microphone or array depends critically on its ability to reject the reverberant sound and on the relative levels of reverberant and direct sounds in the room. This section discusses some of the factors involved.

2.3.1. Single microphone response

The rejection of reverberant sound by a single microphone depends on its "directivity", D_m , defined as the ratio of the response of an omnidirectional microphone to an isotropic sound-field, to the response of the test microphone to the same sound-field:

$$D_m = \frac{[f(0,0)]^2}{\frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi [f(\theta, \phi)]^2 \sin \theta d\theta d\phi} \quad \dots (3)$$

where $f(\theta, \phi)$ is the far-field amplitude/phase response of the microphone in the direction θ, ϕ in spherical polar coordinates. (0, 0) is the direction of the main beam (pointing at the wanted source).

The polar response of any first order microphone (cardioid type) is

$$f(\theta, \phi) = \frac{a + \cos \theta \sin \phi}{1 + a}$$

from which it can be deduced that the directivity is

$$D_m = \frac{(1+a)^2}{a^2 + \frac{1}{3}} \quad \dots (4)$$

Table 1 shows the resultant attenuation of reverberation for various cardioid-type microphones. The maximum attenuation for this type of microphone is 6.0 dB, given when $a = \frac{1}{3}$.

Table 1 – Attenuation Factor

a	Microphone type	Attenuation (dB)
0	Figure-of-eight	4.8
$1/\sqrt{2}$	Hypercardioid	5.4
1	Cardioid	4.8
—	Gun microphone (CK8)	7.8

Gun microphones have especially directional characteristics. One example is the AKG C451 + CK8 adaptor and its characteristic may be approximated to by the forward lobe only of a figure-of-eight. (This is in fact more directional than the actual characteristic.) With this approximation,

the attenuation factor is found to be 7.8 dB, i.e. only 3 dB better than the cardioid microphone. Thus, if the reverberant field is dominant, a gun microphone only gives a slight improvement over conventional microphones.

2.3.2. Array response

The directivity of an array can similarly be computed from (2) and (3) but it is a laborious task since the sources are in the near-field of the array and the 3-D array response is complex. The directivity can, however, be obtained more simply at the low and high frequency limits.

- If the frequency is so low that the wavelength is much greater than the largest dimension of the array (normally, the aperture), then whatever the direction of the wavefront, it will have the same phase at each microphone. The response will therefore be omnidirectional and the directivity unity, resulting in zero rejection of reverberation.
- At high frequencies where the wavelength is much less than the inter-element spacing, the relative phases of the microphone contributions are random³ and the array output is their uncorrelated sum. However, the array response to the direct sound is the correlated sum. Therefore, if equal weighting is given to all N microphones of an array, a net attenuation of $\sqrt{N}$ (i.e. $10 \log_{10} N$ dB's) is given by comparison with a single omnidirectional microphone.

In between the low and high frequency limits, a smooth transition can be expected if the reverberant field is isotropic.

2.3.3. Distance of equal response

To find the overall response of the receiver, it is necessary to determine the relative levels of the direct and reverberant sounds. This is characterised by the "distance of equal response", $\hat{r}_s$, which is the distance from the source at which the direct and reverberant sound levels are the same. It depends not only on the characteristics of the room but also on the directivity of the source, D_s . D_s is defined in a similar way to the microphone directivity, (3), except that $f(\theta, \phi)$ is replaced by the far-field sound pressure level due to the source in free field conditions in the direction θ, ϕ . It is shown in the Appendix that

$$\hat{r}_s = \sqrt{\frac{RD_s}{16\pi}} \quad \dots (5)$$

where R = room constant.

	Large TV Studio	Medium TV Studio	Large Radio Drama Studio	Small Radio Talks Studio	"Small TV Studio"
Volume (m ³)	7257	1968	699	79	1134
Surface area (m ²)	2732	995	516	115	702
Reverberation time (sec.)	0.81	0.31	0.44	0.20	0.74
Average absorption coefficient, $\bar{\alpha}$	0.41	0.60	0.39	0.43	0.30
Room constant, R	1900	1476	331	85	295
$\hat{r}_r$ for omni-directional microphone (m)	6.1	5.4	2.6	1.3	2.4
$\hat{r}_r$ for cardioid microphone (m)	10.6	9.4	4.5	2.3	4.2
$\hat{r}_r$ for 64 microphone array (m)	48.8	43.2	20.8	10.8	19.2

Table 2 : $\hat{r}_r$ for various recording areas

Combining the directivities of the microphone and source, we find the equivalent distance at which the microphone (or array) responds equally to the direct and reverberant sounds:

$$\hat{r}_r = \sqrt{\frac{RD_s D_m}{16\pi}} \quad \dots (6)$$

Table 2 shows the values of R and $\hat{r}_r$ estimated for an omnidirectional source in a variety of recording areas ranging from a small talks studio to a large television studio. The average absorption coefficient $\bar{\alpha}$ is estimated using expression 1 of the Appendix.

For an omnidirectional microphone ($D_m = 1$), $\hat{r}_r$ varies from 1.3 m in the talks studio to 6.1 m in the large television studio. For directional microphones and arrays, D_m is greater than unity and, in the case of a 64-element array, $\hat{r}_r$ is increased by a factor of 8 at high frequencies.

2.4. The combined direct and reverberant sounds

The overall performance of a microphone or array, when both the direct and reverberant fields are accounted for, can be divided into the ability of the array to add "presence", and the ability to discriminate between wanted and unwanted sources. This Section gives theoretical estimates of these parameters.

2.4.1. Presence

The presence of a microphone output signal can be described by its selectivity, the ratio of the direct to reverberant components. If the wanted sound

source is on beam and the array response to the direct sound is defined as unity, then the selectivity (see section 2.3) is

$$10 \log_{10} \left(\frac{\hat{r}_r}{r} \right)^2 \text{ dB for a microphone} \quad \dots (7)$$

and

$$10 \log_{10} \left(\frac{\hat{r}_s}{r} \right)^2 + 10 \log_{10} N \text{ dB for an array} \quad \dots (8)$$

since $\hat{r}_r = \hat{r}_s \sqrt{N}$ for an array at high frequencies.

The first term of (8) describes the relative levels of the direct and reverberant sounds at the array where

$$\frac{1}{r} = \sum_i \frac{1}{r_i}$$

The second term corresponds to the attenuation of reverberation by the array.

2.4.2. Discrimination between sources

The ability of an array to discriminate between two sources, 1 and 2, depends on the response to the direct and reverberant sounds of both sources. If the directivity of both sources is D_s , then the array response (voltage) to source 1 (assumed to be on beam) is

$$\propto \left(\frac{1}{r_1^2} + \frac{1}{N \hat{r}_s^2} \right)^{1/2}$$

and the response to source 2 is

$$\propto \left(\frac{l^2}{r_2^2} + \frac{1}{N\hat{r}_s^2} \right)^{1/2}$$

where l = attenuation of the direct sound from source 2 due to the array beam shape. In each case, the first term is due to the direct sound and the second term to the reverberant sound. The ratio of the response to source 1 to the response to source 2 is then

$$\left[\frac{1 + \frac{1}{N} \left(\frac{r_1}{\hat{r}_s} \right)^2}{l^2 \left(\frac{r_1}{r_2} \right)^2 + \frac{1}{N} \left(\frac{r_1}{\hat{r}_s} \right)^2} \right]^{1/2} \quad \dots (9)$$

The dominant terms in (9) depend on the characteristics of the array and the area in which it is being used. When the reverberant terms dominate, (9) reduces to

$$\left[1 + N \left(\frac{\hat{r}_s}{r_1} \right)^2 \right]^{1/2} \quad \dots (10)$$

which is independent of the location of the unwanted source. When the direct sound dominates, (9) reduces to

$$\frac{r_2}{lr_1} \quad \dots (11)$$

and corresponds to the direct beam attenuation only.

3. Computer predictions

Computer simulations were undertaken to predict the beam shape for various array configurations in response to the direct sound only.

3.1. Design of arrays

The main factors determining the beam shape of an array are its aperture, the number and configuration of the elements and the frequency of the sound source.

If a regularly spaced array is to work over a range of frequencies, its overall size must be greater than a wavelength at the lowest frequency. At the same time, the inter-element spacing must be less than a wavelength at the highest operating frequency if grating lobes are to be avoided. Thus to achieve a good resolution over a wide bandwidth, one configuration might consist of a large array for good low frequency resolution, and a constant microphone spacing corresponding to the smallest value

of λ . This would, however, require some 10,000 elements and is impractical.

One approach, that reduces the number of elements and still achieves a reasonable bandwidth, is to space the microphones logarithmically so that the minimum spacing corresponds to the highest frequency and the maximum spacing to the lowest. Reducing the number of microphones in this way inevitably impairs the performance and it is clear that a compromise has to be sought. The principal considerations are as follows:

1. A large array aperture is needed to give good resolution at low frequencies.
2. The elements should be spaced so that they conform to the Nyquist criterion. To cover a broad band, this implies that some must be closely spaced and others widely spaced.
3. The disposition of the array elements requires careful design to operate over a broad band. A very large number of microphones in the array cannot be contemplated for reasons of cost and practicality. However, a 10×10 element array, configured logarithmically according to condition 2, does not have a large aperture and the resolution at low frequencies is poor. Thus, a compromise has to be sought between the frequency coverage, the array aperture and the number of elements.
4. The distance of the array from the source plane must be chosen carefully. The array should be close enough to the source plane for good selectivity and, at the same time, sufficiently distant that a reasonable area is covered, as determined by sources at glancing angles to the array.

3.2. Array configurations

A large number of array configurations was simulated including linear and log arrays. Of these, only the results of an 8×8 square, log array suspended in the horizontal plane are described here. The array (see Fig. 1) was symmetrical and the inter-element spacing increased logarithmically from 0.1 m at the centre to 1.0 m at the edges. The height of the array above the sound sources was chosen to be 2 m, this figure being a compromise as discussed in the last section. The array elements were omnidirectional at all frequencies, because in practice, this type of microphone has the most consistent gain and phase characteristics.

For each simulation, the position of the wanted source was specified and the response along a line of source positions was measured. The measured positions were in a horizontal plane 2 m below the plane of the array.

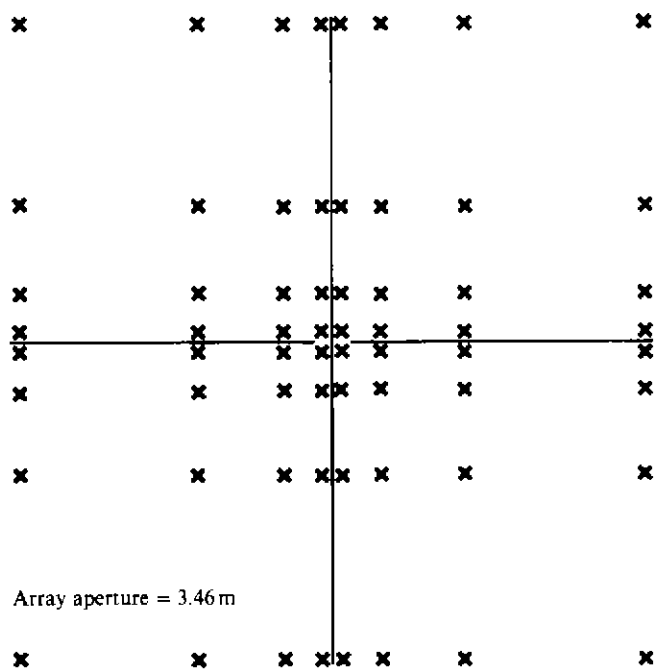


Fig. 1 – Microphone positions of the 2-D logarithmic array.

3.3. Simulation results

The computer simulations showed that regularly spaced, narrow-band arrays can give a narrow beam and good side-lobe suppression at the design frequency. When the same microphones were rearranged logarithmically to cater for a broad band of frequencies, beam broadening occurred and the level of side-lobes increased.

The predicted response of the log array at 1 kHz and 200 Hz is shown in Fig. 2. Co-phasing and Hanning weighting were included. At 1 kHz, a 1 m

wide beam was formed with side-lobes attenuated by more than 10 dB. At frequencies below 200 Hz, the array was less directional than a single figure-of-eight microphone at the same distance. For off-axis sources, the beam was much less directional than for a source at the broadside position. Fig. 3 shows the beam-shape of the array when the intended source position was 45° off-axis.

In the wanted source direction, the frequency response of the array was flat, but in other directions it was similar to that of a comb-filter with a falling frequency response. Similar but less drastic effects occur with single directional microphones.

The tolerance to phase and gain errors was better than expected. Phase errors of 50° and gain errors of 5 dB altered the beam shape, but did not seriously degrade it. In practice, the phase and gain characteristics of omnidirectional microphones should meet these figures at speech frequencies. However, serious difficulties could be encountered at high frequencies and in cases where high attenuation is wanted.

Directional microphones have the potential to improve the selectivity of the array. A detailed investigation of the phase and gain integrity of such microphones was not undertaken, but a cursory examination was not promising. At high frequencies, the variation with azimuth of the gain and especially the phase of directional microphones can be considerable. Also, sources in the near-field of an array would be at different azimuths relative to the axis of each microphone. Thus, although relatively relaxed tolerances are allowed, directional microphones could present difficulties in practice and they were not investigated further.

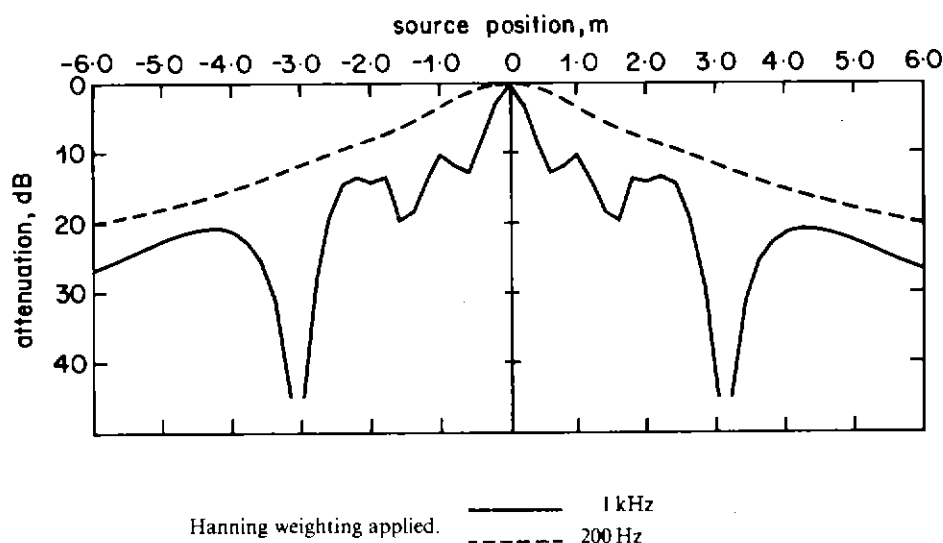


Fig. 2 – Predicted response of the 2-D log array, on axis.

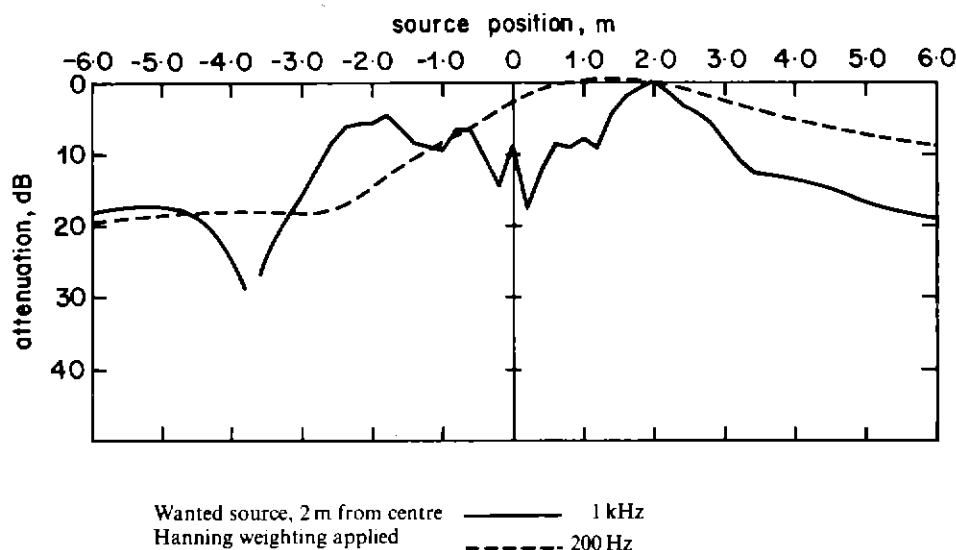


Fig. 3 – Predicted response of the 2-D log array, off axis.

To summarise, the log array described gave good beam-shapes at mid and high frequencies for on-axis sources, but at low frequencies, it was hardly an improvement on a single figure-of-eight microphone at the same distance. Although some improvement could be made by reconfiguring the array, perhaps by increasing its overall size, it seems unlikely that the improvement would be very significant.

A technique that could give some improvement is to equalise the element signals according to their position. At low frequencies, the array would appear large and the elements widely spaced, whilst at high frequencies it would appear small and the elements closely spaced. This might help to avoid the effects of aliasing sometimes apparent in the computer plots. Although Hanning weighting was beneficial at high frequencies, it slightly impaired the beam-shape at low frequencies. Some form of frequency compensation might also be fruitful here. Other types of weighting (see, for example, Refs. 1 and 2) might also be beneficial. These techniques were not, however, explored.

4. Practical tests

4.1. Array Design

Practical tests, using simple 8-element arc arrays, were performed to test the validity of the theoretical models, to investigate the practicalities of setting up and using arrays and to allow subjective appraisal of arrays on speech. Arc arrays, where the microphones are equidistant from the wanted source, were used for the experiments in order to simplify the co-phasing requirement. Computer plots of beam-shapes were made to help choose the most suitable arrays for testing.

Several arrays were tested and the first (shown in Figs. 4 and 5) was predicted to give side-lobes as low as -30 dB. This degree of attenuation was an important test of the arrays, since it explored the maximum selectivity which could be obtained with practical tolerances. The array was regularly spaced with an inter-element spacing of 0.4 m resulting in an aperture of 2.8 m and giving the best beam-shape in the 500 – 1000 Hz band. It was also about the largest array that would fit in the free-field room used for the tests, thereby minimising the relative sensitivity of the array to positional errors.

A source-to-array distance of 3 m was chosen. At this distance, the source was not too deep into the near-field of the array and the array did not need to be sharply curved to achieve co-phasing.

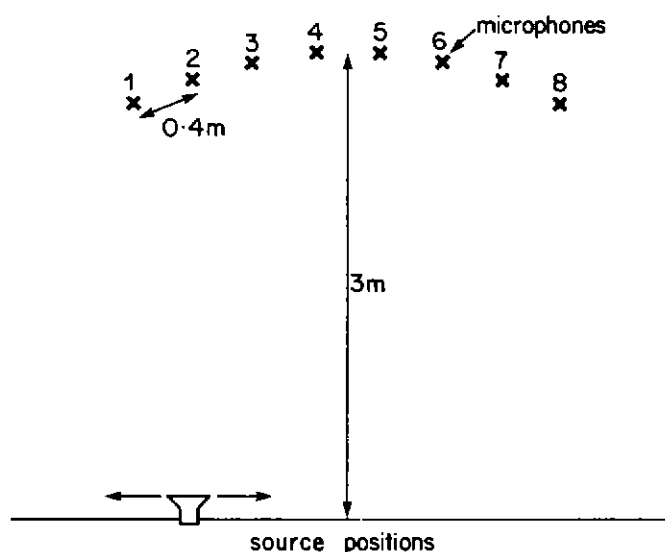


Fig. 4 – Test arrangement for the 0.4 m spaced arc array.

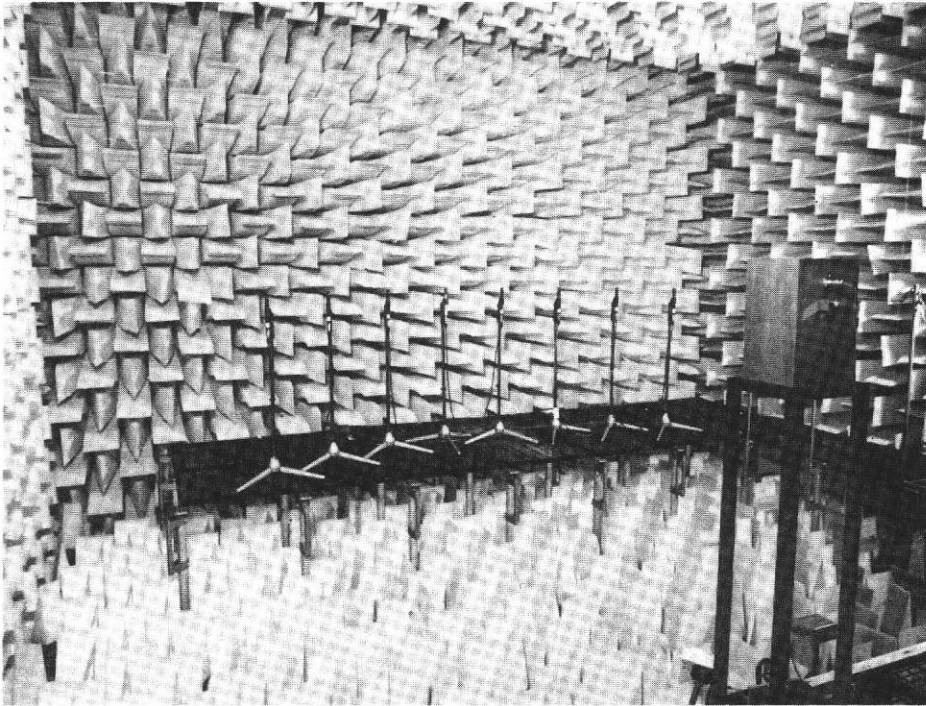


Fig. 5 – View of the 0.4 m spaced arc array in the Free Field Room.

At high frequencies (above 4 kHz), computer plots of this array's beam-shape showed a lot of fine detail and closely spaced grating lobes. They arose because the microphone spacing was large compared with the wavelength at these frequencies. An array of one quarter the size (0.1 m spacing, 0.7 m aperture) was therefore chosen for testing the performance at high frequencies.

The third array tested had a microphone spacing increasing logarithmically from the centre of the array and was chosen as an example of a broadband array. The spacing increased in powers of 2 from 0.1 m to 0.8 m, resulting in an overall aperture of 2.9 m.

4.2. Experimental arrangements

Three different areas were used to test the arrays; a free-field room to measure the response to direct sound only, a reverberation room to measure the response to reverberant sound only, and a room similar in size and acoustic to a small television studio.

The arrays each consisted of eight AKG CK22/C451 omnidirectional microphones connected to an 8-channel mixing desk. A LS 3/5A monitoring loudspeaker was used as the movable sound source except for speech tests when a live speaker was used. For the free-field room tests, pure tones were emitted from the loudspeaker, whilst for the reverberation room and the studio tests, white noise was emitted to average the room effects.

The microphones and loudspeakers were set up on stands, great care being taken to position the microphones accurately. It was found possible to position them on an arc centred on the loudspeaker to a tolerance of ± 3 mm.

The gains of the microphones were set up by adjusting the desk faders for equal levels (or the appropriate Hanning weighting) from each microphone when the loudspeaker was in the central position emitting a 500 Hz tone. The loudspeaker was moved along a line perpendicular to the axis of each array (see Fig. 4) and plots made of each array's response.

4.3. Component amplitude and phase responses

The amplitude and phase responses of the mixing desk channels, microphones and loudspeaker were checked to ensure that they would not degrade the array's performance. Since the experimental tolerances achieved in setting up the microphone positions and gains were around ± 3 mm (equivalent to 3 degrees of phase at 1 kHz) and ± 0.5 dB respectively, it was desirable that the response of the microphones, loudspeaker and mixing desk should lie well within these limits so as not to cause further deviations from prediction.

The results of measurements on the mixing desk, microphones and loudspeaker were as follows:

- a) The mixing desk channels met the limits easily. The amplitude response of each channel was

flat to within ± 0.5 dB from 20 Hz to 20 kHz. Phase differences between the channels were within 1° at 1 kHz rising to a maximum of 3° at 10 kHz.

- b) The polar amplitude and phase responses of the microphones were checked up to 2 kHz. Each microphone under test was clamped vertically 1 m from a loudspeaker emitting pure tone in the free-field room. The amplitude and phase of each microphone output was measured as the microphone was turned in the clamp. The amplitude responses were found to be flat with azimuth to within ± 0.2 dB and the phase responses were flat to within $\pm 1^\circ$. The microphones were also checked for mutual consistency of phase response up to 2 kHz. The phases were within $\pm 2^\circ$ at 1 kHz and $\pm 7^\circ$ at 2 kHz. Some of the spread could have been caused by slight shifts of the microphone positions in the clamp.
- c) The loudspeaker's polar response was measured using a microphone on a rotating boom centred on the loudspeaker. It was found that the apparent centre of radiation of the loudspeaker moved back by about 100 mm as the frequency was increased from 250 Hz to 2 kHz. Allowing for this, the phase variation was constant within $\pm 3^\circ$, at any single frequency up to 2 kHz within the azimuths $\pm 45^\circ$ (the maximum angles subtended by the arrays at the source position). The polar amplitude variation of the loudspeaker, measured using the same arrange-

ment, was less than ± 0.5 dB within the azimuths $\pm 45^\circ$ up to 500 Hz. Above this frequency, there was a fall-off in amplitude with azimuth reaching -4 dB at 45° at 2 kHz. This fall-off was equivalent to an extended source, 130 mm wide, giving rise to a sinc x diffraction pattern. It had only a slight effect on the measurements of the array beam-shape.

5. Results of the practical tests

5.1. Direct sound in free-field room

The microphones and loudspeaker were set up as described in Section 4.1 on a steel grid floor in the free field room. As little floor as possible was laid to minimise reflections.

5.1.1. Tone tests

Fig. 6 shows the response of the 0.4 m regularly spaced array without Hanning weighting at 500 Hz compared with the computer prediction. There is agreement to within ± 1 dB over most of the plot and the positions of the nulls are accurately reproduced. The deviation from prediction lies within that caused by experimental errors.

Further plots were made in octave steps from 250 Hz to 4 kHz for the 0.4 m spaced array and grating lobes (see Section 2.2) appeared at frequencies above 1 kHz. Deviations from the predicted

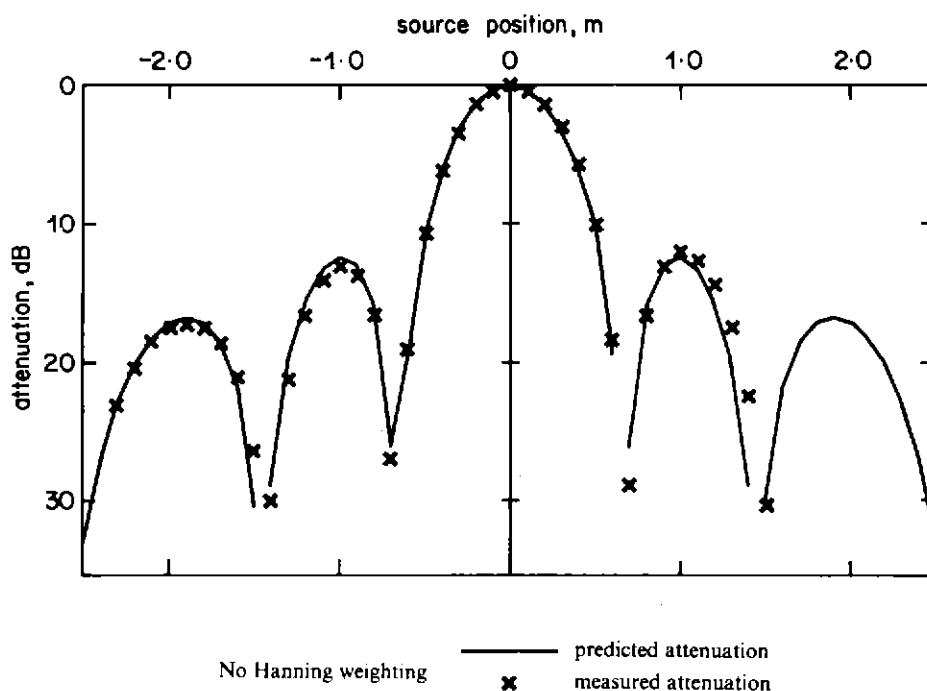


Fig. 6 – 0.4 m spaced arc array at 500 Hz in the FFR.

beam-shape again lay within the experimental error due to positioning.

The 0.1 m spaced array was tested at various frequencies. At 8 kHz the sidelobes rose to about 8 dB higher than predicted (see Fig. 7). This deviation was greater than that predicted for the experimental tolerances. A separate experiment revealed phase perturbations at the microphones above about 3 kHz, attributable to reflections of sound off the microphone assembly and steel grid floor. They probably account for the discrepancy.

The beam-shape of the logarithmically spaced array at 1 kHz is shown in Fig. 8. The correspondence with prediction is good. It can be seen that the level of sidelobes (-8 dB) is higher than that for the regular spaced array (-12 dB).

To summarise, the measured array responses to the direct sound were found to be close to prediction. The measured peak level of the sidelobes matched the predicted levels up to 1 kHz, but rose above the predicted level by about 5 dB at 2 kHz and 8 dB at 8 kHz. Up to 2 kHz, the deviations from prediction lay within the expected limits of error. At higher frequencies, however, the deviations exceeded expectations and were probably caused by sound reflections off the arrays' supporting structure.

These results therefore can be taken to confirm the validity of the computer predictions.

5.1.2. Speech tests

Live speech tests were carried out with all three arrays. The speaker read a passage of test from positions at 25 cm intervals along the line of source positions (see Fig. 4). At each position, the speaker read the passage once facing straight ahead, and a second time turning his head rapidly from side to side to show the effects of shifting to face different directions. The results were recorded on tape together with the output from one of the omnidirectional microphones (No. 4 in Fig. 4) for comparative purposes.

The first test used the 0.4 m regularly spaced narrow band array with Hanning weighting. The level of the speech was heard to fall as the speaker moved off-axis. On axis, good sound quality was given which was not immediately distinguishable from the single microphone. Head movement produced an unpleasant "swishing" effect. As the speaker moved off-axis, the voice suffered severe tonal changes due to the comb-filtering effect of the array. This was evident on a spectrum analyser as a band of attenuation moving down with frequency as the speaker moved off axis. The lower edge of the

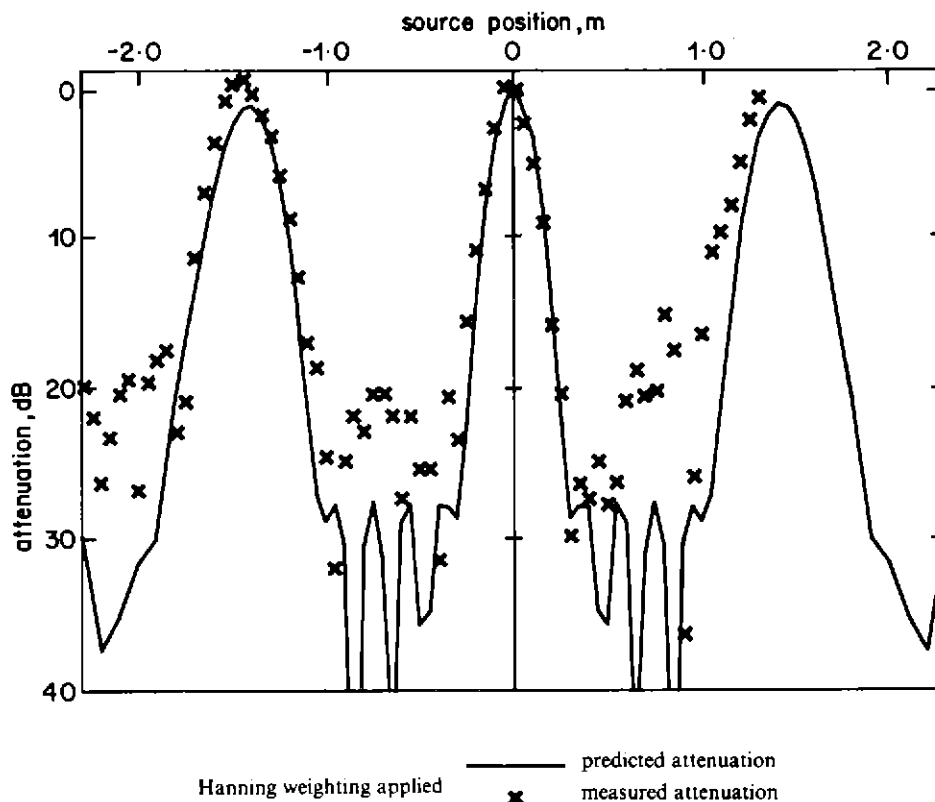


Fig. 7 – 0.1 m spaced arc array at 8 kHz in the FFR.

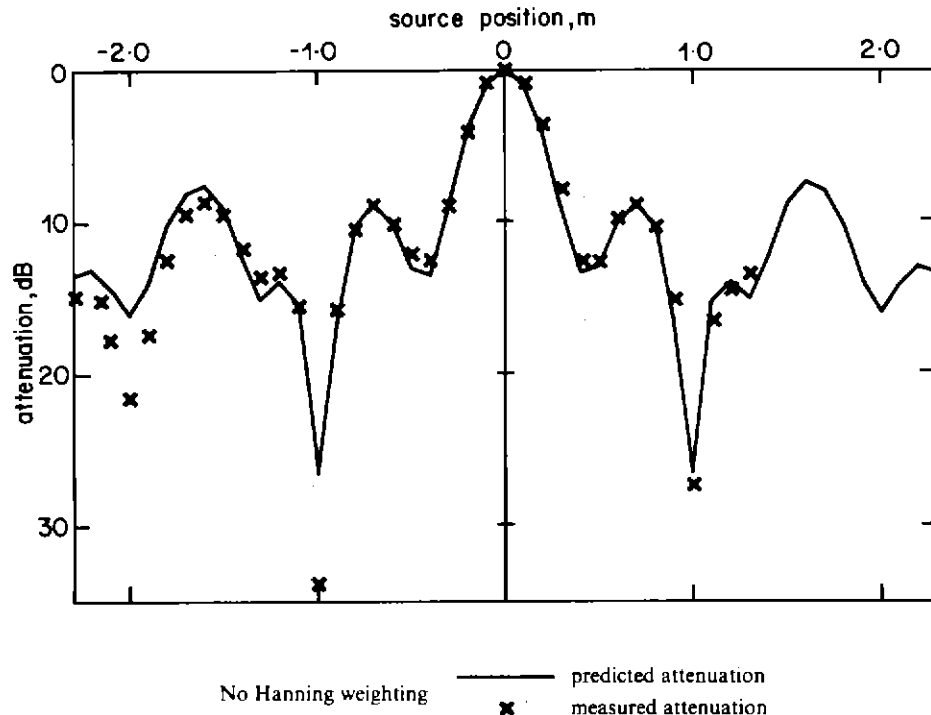


Fig. 8 – Log array at 1 kHz in the FFR.

band corresponded to the array's beam-width being reached at progressively lower frequencies. The upper edge corresponded to the appearance of the first grating lobes.

The logarithmically spaced array also gave good sound quality on axis. Some swishing could be heard with head movement. There was again coloration of the sound as the speaker moved off axis, but it was not as bad as with the regularly spaced array.

5.1.3. Microphone noise

Microphone noise was very noticeable on the output of the array, although the measured signal-to-noise ratio from the array was better by 4.8 dB than that from the single microphone. The improvement was due to the lack of correlation of the noise signals from each microphone giving a net attenuation of $\sqrt{N}$ relative to the single microphone. For an eight microphone array, the S/N ratio improvement should be 9 dB, ($20 \log_{10} \sqrt{8}$ dB). The array used in the free field room had Hanning weighting applied, which brought the theoretical improvement down to 7.9 dB. However, in practice, an improvement of only 4.8 dB was measured because one of the microphone channels was about 6 dB noisier than the others.

The reason why the S/N ratio was so poor compared with a normal speech recording was simply that the microphones were further from the

source than is normal practice. A single microphone would normally be located some 60 cm from a source whereas the array was 3 m distant, thereby causing a loss of 14 dB of gain. The gain increase due to the coherent addition of the element contributions only improved this figure by the above mentioned 5 dB and a poor overall S/N ratio resulted. Thus, in general, the improvement in signal-to-noise ratio of an array may not compensate for the drop in signal level caused by the extra distance from the sound source.

5.2. Reverberant sound in reverberation room

5.2.1. Noise tests

The arrays were set up individually in the middle of the reverberation room. A diffuse reverberant field was generated by pointing a loudspeaker into one corner of the room and emitting white noise from it. The outputs of the array, and of each microphone individually, were measured in 1/3-octave bands with a spectrum analyser. A 1-minute integration time gave sufficient time averaging to obtain a reproducible reading in each band.

To check whether any room effects were interfering with the true response of the array, the measurements were made on both the 0.4 m and the 0.1 m regularly spaced arrays. The response of the 0.1 m spaced array was expected to be the same as the response of the 0.4 m spaced array shifted up in frequency by 2 octaves.

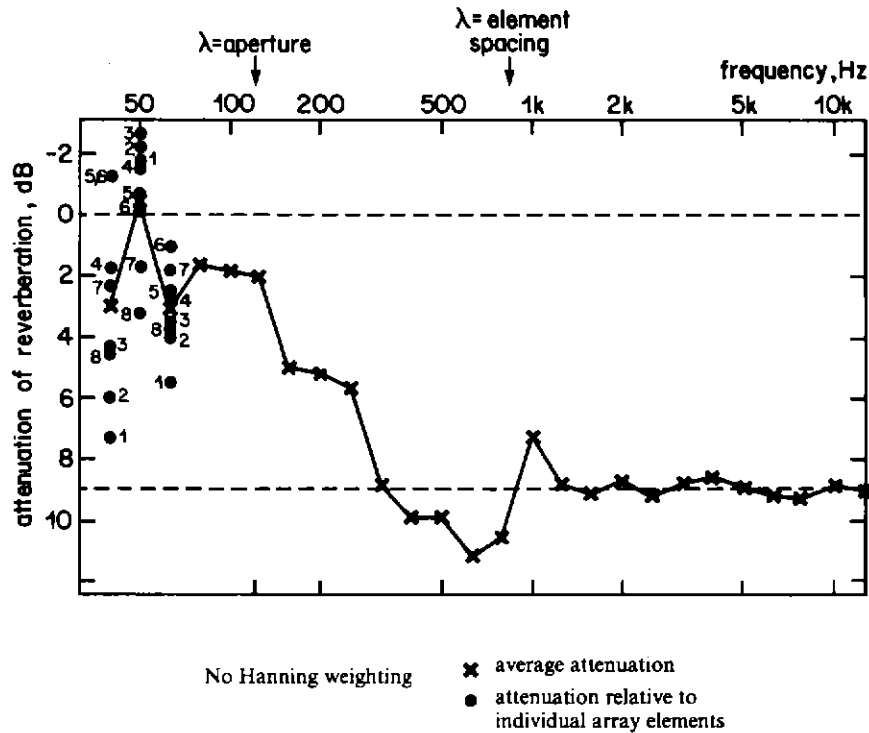


Fig. 9 - 0.4 m spaced arc array in the reverberation room.

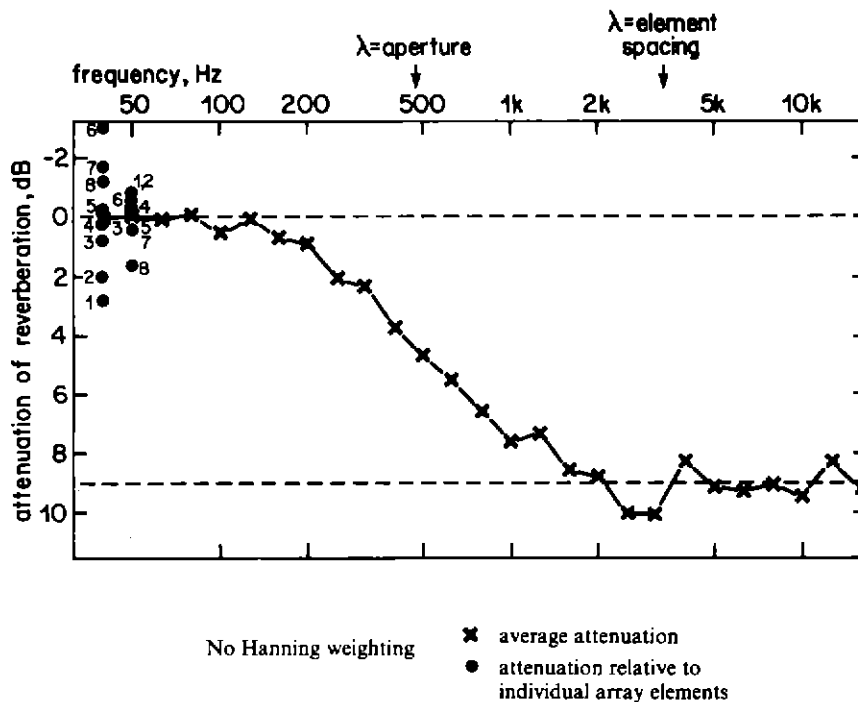


Fig. 10 - 0.1 m spaced arc array in the reverberation room.

Figs. 9 and 10 show the rejection of reverberation vs. frequency by the 0.4 m and 0.1 m regularly spaced arrays respectively. The rejection was found by measuring the difference in level between the array output and a single microphone output, and inserting an 18 dB offset corresponding to the gain of

the array. Both plots show a transition from zero rejection at low frequencies to a high frequency limit of 9 dB as predicted. The high frequency limit of rejection is reached when the wavelength is about twice the inter-element spacing; this is fortunate since this is the region in which the array would give

its optimum beam-shape. The low frequency limit is reached when the wavelength is approximately four times the aperture of the array.

As well as plotting the average result, the array output was plotted at low frequencies relative to each individual microphone output (numbered as shown in Fig. 4). In the lowest frequency bands, the sound field was not isotropic since the wavelength was comparable with the size of the room and few room-modes existed. Standing waves can be seen superimposed on the array in bands centred on 40, 50 and 63 Hz; following the order of the microphone numbers, the sound level can be seen to rise and fall.

5.2.2. Speech tests

A speech test was made in the reverberation room with the 0.4 m regularly spaced array using the same procedure as for the free field room tests. The array showed no discrimination at all due to the overwhelming level of reverberant sound. When compared with a single microphone, the output of the array could be heard to have a slight bass-rise due to the 9 dB lift in response at low frequencies.

5.3. Performance in a "small TV studio"

Two of the arrays tested in the free field room, the 0.4 m regularly spaced array and the 0.1 m to 0.8 m logarithmically spaced array, were tested in a large room similar in size and acoustic to a small

television studio*. The arrays were set up in the middle of the room with the microphones 1 m above the floor. Plots of beam-shapes were made in the same way as in the free field room, but using 1/3-octave noise as the test sound source in order to avoid problems caused by standing waves. Speech tests were also made with these arrays.

5.3.1. Noise tests

Fig. 11 shows the response of the regularly spaced array (without Hanning weighting) in the "studio" compared with the predicted response of the same array under free field conditions. The studio plot is blurred by about $\pm 12\%$ along the distance axis since 1/3-octave noise was used as the test sound instead of pure tone. However, it can be seen that the attenuation of the sidelobes is limited to about 13 dB and the individual sidelobes merge. This is due to the reverberant sound.

Table 1 gives the predicted distance of equal response for an omnidirectional source and expression (10) gives the limit of an array's discrimination in a room due to reverberation. Polar response measurements of the loudspeaker showed that its directivity is approximately 2 (3 dB) at 500 Hz; $\hat{r}_s$ is then estimated to be 3.4 m in the

*The dimensions of the room were 18 m (length) $\times$ 9 m (width) $\times$ 7 m (height) and its average reverberation time in the band 250 Hz–4 kHz was 0.7 seconds.

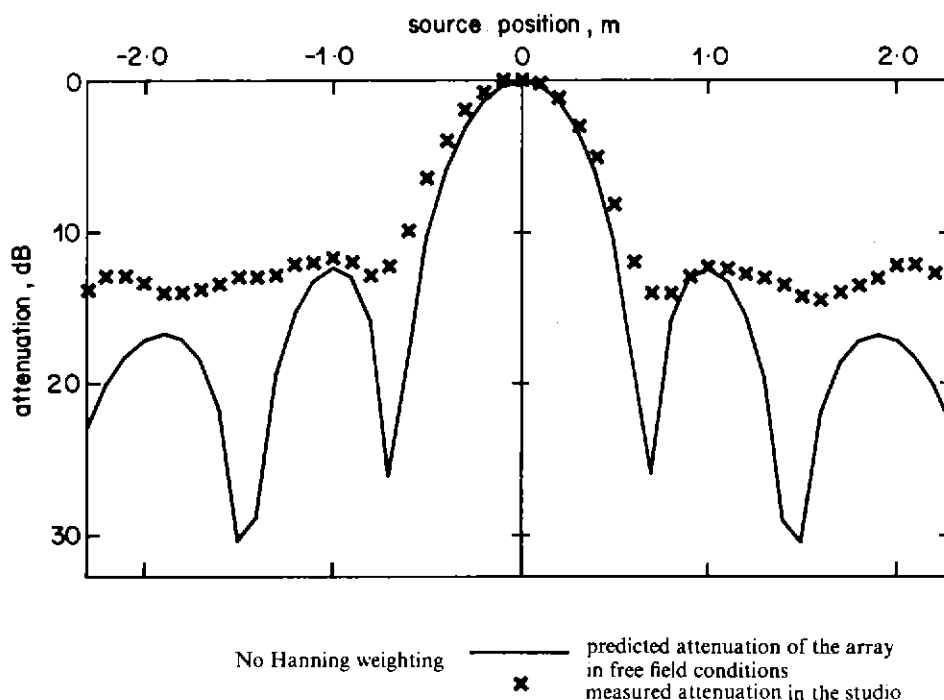


Fig. 11 – 0.4 m spaced arc array at 500 Hz in the studio.

“studio”. The resultant limit of discrimination deduced from (10) is 10.5 dB. This is within 2.5 dB of the measured figure. Given the assumptions used to derive (10), the predicted limit of the array’s discrimination due to reverberation is reasonably close to the measured value. The residual discrepancy gives some indication of the errors in the assumptions made of the prevailing conditions, especially of the diffuseness of the reverberant field.

5.3.2. Speech tests

Speech tests were carried out with the regularly spaced and the logarithmically spaced arrays, using the same procedure as for the free field room tests. The two arrays had a similar performance. On axis, the speaker was noticeably “brought forward” compared with the single microphone because of the attenuation of reverberation. Off axis, the speaker sounded more distant and “echoey” as the direct sound was rejected, leaving only the reverberant sound. In both cases, the reverberation could be heard to be the limiting factor and remained at a constant level as the speaker moved further off axis. The reverberation also filled in the nulls in the comb-filter frequency response off axis, thereby giving a more natural sound than in the free field room. However, the tonal quality was coloured by the high levels of reverberation at low frequencies. The subjective attenuation of the speech by both arrays was roughly 6 dB at 1 metre off axis.

All the speech recordings made in the studio suffered from a high level of background noise. A loud, low frequency rumble from nearby plant could be heard on the array output although it was barely audible in the room. This was due to the combined effects of the low gain of the array (due to its distance from the wanted source) and to the poor attenuation of low frequency reverberant sound (see Section 2.3). In practical applications, this could represent a significant problem.

6. Discussion

6.1. Practical tests

The practical tests have largely confirmed the theoretical model given in Section 2. The free-field room tests showed that the direct sound beam-shape accurately matched the computer predictions within experimental error. At high frequencies, some deviation was observed but was attributable to reflections off the grid floor and microphone stands. The response to the reverberant field as measured in the reverberations room also largely met expectations. The attenuation increased from zero at low frequencies to $\sqrt{N}$ at high frequencies. The effects of

a non-diffuse field were evident at low frequencies as standing waves; similar effects are, if anything, more likely in real recording areas despite the acoustic treatment that such areas receive.

The performance of the arrays was limited by the small number of microphones. The attenuation of the reverberant field was small (9 dB max.) and the direct sound beam-shape was therefore masked. The beam forming capability was also poor and effective only over a narrow band of frequencies. Real arrays would require much larger numbers of microphones and some possibilities are discussed in Section 6.2.

Despite the limited performance, valuable subjective and practical characteristics of the arrays were revealed. Speech quality sounded natural on-axis but, particularly in free field conditions, comb-filtering effects coloured the sound. Broad-band arrays, however, should show an improvement in this respect. In reverberant areas, the nulls of the comb-filter were “filled-in” by the reverberant sound and the colorations were not so severe.

Small errors in the positioning of the microphones cause larger phase errors at high frequencies. Thus, although tolerances of ± 3 mm were achieved, residual errors were evident as a slight high frequency attenuation of the wanted (on-beam) sound source, and as a lower attenuation of the side-lobes at high frequencies. Stringent positional tolerances would have to be imposed therefore to achieve good wide band performance with practical arrays.

The loss of gain of the array due to its distance from the source was only partially offset by the increase in gain due to the number of microphones. As a result, there was a net increase of electronic noise in the array output by comparison with a closely positioned spot microphone. However, a practical array is likely to need far more than 8 microphones and it is probable then that the electronic noise would be adequately attenuated.

Tests in the “small studio” showed that the array was sensitive to low frequency, plant noise. This is due to the negligible attenuation of reverberant sound at low frequencies. To improve the rejection, the dimensions of the array would have to be greatly increased beyond those (2.8 m) of the test arrays. This could be impractical in real recording studios. Rejection of low frequency sounds is important since it is difficult to exclude low frequency rumble, caused by ventilation plant, from studio areas. Spot microphones achieve the necessary broadband rejection by their close proximity to the source and hence high gain.

6.2. Characteristics of practical arrays

The theoretical predictions and practical tests have indicated that, for a microphone array to give worthwhile performance for recording purposes, a large number of microphones are needed. This is primarily because arrays require many elements to achieve the same gain as a spot microphone close to the source. Estimates are made in this Section of the number of elements in an array that would give results comparable with a spot cardioid microphone, 0.75 m from a source – typical recording conditions for speech in a television studio.

As indicated earlier two important functions of a microphone are to give presence to the wanted sound and to discriminate against unwanted sounds. Experience with single microphones in television studios indicates that a ratio of the direct to reverberant sound levels should be 25 dB or more to give speech adequate presence. This figure is based on a cardioid microphone, 0.75 m from a source with directivity 2, in a medium sized television studio where $\hat{r}_s = 7.6$ m.

Suppose the array is 3 m from the source and that it is to be used in the same studio as the cardioid microphone. Then, if we make the assumption that the maximum rejection of reverberation is achieved (i.e. $1/\sqrt{N}$), the minimum number of microphones needed would be 49 to achieve the same presence figure. However, the maximum rejection is only achieved when the inter-element spacing $d > \lambda/2$ (see Section 5.2.1). A broadband array would probably require at least double the number of 49 elements.

The discrimination of the cardioid microphone against unwanted sounds was computed using expression (9) under a variety of conditions. If the array is to achieve comparable performance, it would need at least 48 microphones (to attenuate the reverberant sounds) and side-lobes would have to be attenuated by approximately 18 dB. These figures are again optimistic. When broadband requirements and more realistic estimates of the rejection of reverberation are taken into account, then the number of array elements would have to be considerably increased.

The estimates of the performance requirements of the array could be relaxed if some loss of quality were acceptable. Siting the array in a large studio would also ease the requirements since $\hat{r}_s$ is larger and the levels of reverberation are reduced. Nevertheless, a practical array is still likely to need at least 50 microphones. The cost of such an array would also be an important factor. 50 microphones and pre-amplifiers together with the complex digital hardware to steer the array would be expensive.

7. Conclusions

The performance of arrays in studios is complicated by the presence of reverberant sound and the near-field conditions which prevail at most frequencies. Theoretical models describing these conditions have been formulated and confirmed, within experimental accuracy, by practical tests on arc arrays. They can therefore form the basis for designing practical arrays.

A detailed design of a wide-band array was not undertaken, partly because of the complexity required for broadcast quality sound but also because it was apparent that a large number of microphones would be needed. Rough design estimates were made, however, for an array in a medium sized television studio, an area where the reverberant sound levels are relatively low. If the array is to be as selective as a cardioid microphone at 75 cm from a speaker (typical recording conditions), then at least 50 microphones would be needed, although a more realistic figure is over 100. The reason for such high figures is mainly because the array has to match the high selectivity of a single microphone close to the source.

A number of practical difficulties are likely to be encountered in fully designing a practical array. If broadcast quality speech is to be achieved, then the positional tolerances of the microphones are severe since the selectivity of the array is achieved by phase cancellation. A second factor is that reverberant sound and background noise (due to ventilation plant for example) are usually greatest in level at low frequencies. At these frequencies, an array's rejection of reverberation is poor unless the array has a very large aperture (e.g. 12 m). Coverage of an array is a further factor. For maximum gain, the array would be located close to the sources. Off-axis sources then subtend glancing angles at the array and the array beam-shape is poor. Alternatively, the array has to be at some distance from the sources and this places more stringent design requirements for beam forming and rejection of reverberation. Perhaps the most significant difficulty is to achieve good wide-band beam-shapes whilst maintaining good wide-band rejection of reverberation. The sampling criterion for the first is that the inter-element spacing $d < \lambda$, and the condition (found experimentally) for the second is that $d > \lambda/2$. Reconciling these two conditions over a broad band of frequencies requires some ingenuity if the number of microphones is not to be too great.

The considerable design and practical difficulties and costs of arrays are such that it is unlikely that arrays can form a viable alternative as a general

purpose recording technique in studios. However, circumstances may arise where arrays could fulfil a useful role. If some loss of sound quality is acceptable (limited band-width and some increase in reverberant sound level) or if the reverberant sound level is low (e.g. out-of-doors), then the design difficulties are considerably eased and smaller numbers of microphones can be contemplated. One example is a 1.2 m wide portable arc array of 40 electret microphones, designed for speech and special effects applications.⁴

8. References

1. CURTIS, T.E. and WARD, R.J. Digital beam forming for sonar systems. IEE Proc. F, Communications, Radar and Signal Processing, 127, 4, pp. 257-265, August 1980.
2. GRIFFITHS, J.W.R. Adaptive array processing. A tutorial. IEE Proc. F, Communications, Radar and Signal Processing, 130, 1, pp. 3-10, February 1983.
3. KUTTRUFF, H. "Room Acoustics". Applied Science Publishers, 1973, pp. 218-223.
4. ABE, Y. *et al.* Practical application and digital control of the microphone array. 76th AES Convention, New York. Pre-print 2116 (E-1), October 1984.
5. BERANEK, L.L. Acoustics. McGraw-Hill, 1954.

Appendix 1

The distance of equal response

The sound pressure level (spl) due to the direct sound from a source may be expressed in terms of its directivity, D_s . This is defined (see e.g. Ref. 5) as the ratio of the sound intensity at a designated point relative to the source (normally on the axis of maximum radiation), to the corresponding intensity due to an omni-directional source of the same acoustic power as the directional source.

The intensity, I , due to the omni-directional source is independent of direction and is given by

$$W_{\text{omni}} = 4\pi r^2 I$$

where W = acoustic power of source
and r = distance from the source.

The on-axis intensity at a point relative to the directional source is

$$I_{\text{dir}} = \frac{|p(0,0)|^2}{\rho_0 c}$$

where $p(0,0)$ = on-axis sound pressure at a distance, r ,

ρ_0 = air density

and c = speed of sound.

Thus, the source directivity,

$$D_s = \frac{|p(0,0)|^2}{\rho_0 c} \cdot \frac{4\pi r^2}{W_{\text{omni}}}$$

The on-axis spl of the directional source may then be expressed as

$$|p(0,0)|^2 = \frac{W D_s \rho_0 c}{4\pi r^2}$$

One can also show that the spl due to the reverberant sound, if diffuse, is given by

$$p^2 = \frac{4\rho_0 c W}{R}$$

where R , room constant = $\frac{S\bar{\alpha}}{1-\bar{\alpha}}$

$\bar{\alpha}$ = average absorption coefficient of the surface of the enclosure*

and

S = total surface area

Thus, the levels of the direct and reverberant sounds are equal when the distance from the source,

$$r_s = \sqrt{\frac{R D_s}{16\pi}}$$

* $\bar{\alpha}$ may be found from the Norris-Eyring equation for the reverberation time,

$$T_{60} = -\frac{0.161V}{S \ln(1-\bar{\alpha})} \quad \dots (1)$$

where V = volume of the enclosure and all dimensions are in metres.

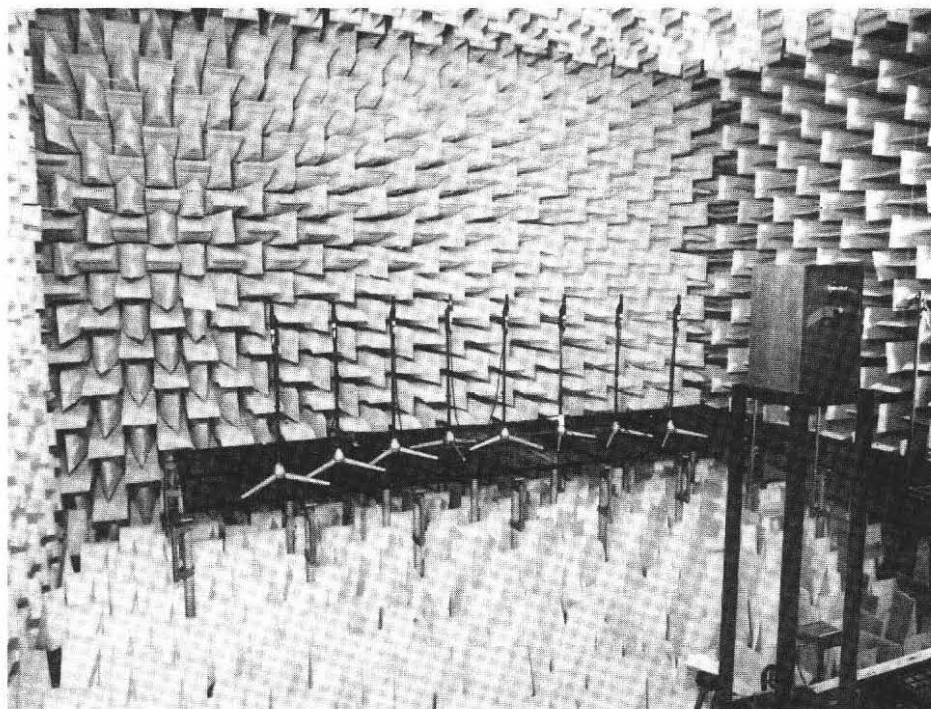


Fig. 5 – View of the 0.4 m spaced arc array in the Free Field Room.

At high frequencies (above 4 kHz), computer plots of this array's beam-shape showed a lot of fine detail and closely spaced grating lobes. They arose because the microphone spacing was large compared with the wavelength at these frequencies. An array of one quarter the size (0.1 m spacing, 0.7 m aperture) was therefore chosen for testing the performance at high frequencies.

The third array tested had a microphone spacing increasing logarithmically from the centre of the array and was chosen as an example of a broadband array. The spacing increased in powers of 2 from 0.1 m to 0.8 m, resulting in an overall aperture of 2.9 m.

4.2. Experimental arrangements

Three different areas were used to test the arrays; a free-field room to measure the response to direct sound only, a reverberation room to measure the response to reverberant sound only, and a room similar in size and acoustic to a small television studio.

The arrays each consisted of eight AKG CK22/C451 omnidirectional microphones connected to an 8-channel mixing desk. A LS 3/5A monitoring loudspeaker was used as the movable sound source except for speech tests when a live speaker was used. For the free-field room tests, pure tones were emitted from the loudspeaker, whilst for the reverberation room and the studio tests, white noise was emitted to average the room effects.

The microphones and loudspeakers were set up on stands, great care being taken to position the microphones accurately. It was found possible to position them on an arc centred on the loudspeaker to a tolerance of ± 3 mm.

The gains of the microphones were set up by adjusting the desk faders for equal levels (or the appropriate Hanning weighting) from each microphone when the loudspeaker was in the central position emitting a 500 Hz tone. The loudspeaker was moved along a line perpendicular to the axis of each array (see Fig. 4) and plots made of each array's response.

4.3. Component amplitude and phase responses

The amplitude and phase responses of the mixing desk channels, microphones and loudspeaker were checked to ensure that they would not degrade the array's performance. Since the experimental tolerances achieved in setting up the microphone positions and gains were around ± 3 mm (equivalent to 3 degrees of phase at 1 kHz) and ± 0.5 dB respectively, it was desirable that the response of the microphones, loudspeaker and mixing desk should lie well within these limits so as not to cause further deviations from prediction.

The results of measurements on the mixing desk, microphones and loudspeaker were as follows:

- a) The mixing desk channels met the limits easily. The amplitude response of each channel was